LINEAR ALGEBRA AND ITS APPLICATIONS

ELSEVIER Linear Algebra and its Applications 284 (1998) 125–135

# On-line change-point detection (for state space models) using multi-process Kalman filters

Martin Daumer *, Markus Falk

*Institute for Medical Statistics and Epidemiology, Klinikum rechts der Isar, Technical University Munich, Ismaninger Str. 22, 81675 Munich, Germany*

Received 11 August 1997; accepted 29 June 1998

Submitted by P.N. Shivakumar

**Abstract**

In this paper we show how to use an on-line algorithm based on a multi-process-Kalman-filter–extending ideas described in Whittaker and Frühwirth-Schnatter (J. Whittaker, Frühwirth-Schnatter, Appl. Stat. 43 (4) (1994)) – to detect sequential change-points in noisy time series. We focus on types of change-points typically arising in biomedical signals, i.e. jumps or drifts in nonstationary time series possibly corrupted by embedded outliers. The algorithm has been implemented in a program written in Matlab 5.0 and was tested using vital parameters recorded during surgical procedures performed at the University Hospital of the Technical University of Munich, Klinikum Rechts der Isar, Munich. 

## 1. Introduction

On-line monitoring of time series is becoming increasingly important in different areas of application such as medicine, biometry and finance. In medicine, on-line monitoring of patients after kidney transplantation [1] is a prominent example. In finance, fast and reliable recognition of changes in level and trend of intra-daily stock market prices is of obvious interest for ordering

* Corresponding author. E-mail: martin.daumer@imse.med.tu-muenchen.de


PII: S0024-3795(98)10132-5

and purchasing. In this project, we currently consider monitoring of surgical data like heart rate, blood pressure and oxygenation.

Change-points in such time series have to be detected in real time as new observations come in, usually in short time intervals. Retrospective detection of change-points, after the whole batch of observations has been recorded, is nice but useless in monitoring patients during an operation.

There are various statistical approaches, particularly in statistical quality control, which are conceivable for on-line detection of change-points in time series. Recently dynamic models or state space models have been discussed for autocorrelated data, and they seem particularly well suited for our purposes. "Filtering" was historically developed exactly for on-line estimation of the "state" of some system. Our approach is based on an extension of the so-called multi-process Kalman filter for change-point detection [2]. It turns out, however, that some important issues for an adequate and reliable application have to be considered, in particular the (appropriate) handling of outliers and, as a central point, adaptive on-line estimation of control- or hyper-parameters. In this paper we describe a filter model that has these features and can be implemented in such a way that it is useful for real time applications with high sampling rates.

Recently, simulation based methods for estimation of non-Gaussian dynamic models have been proposed that may also be adapted and generalized for the purpose of change-point detection. Most of them solve the smoothing problem, but very recently some proposals have been made that could also be useful for filtering and, thus, for on-line monitoring [4–6]. Whether or not these approaches provide a useful alternative to our development requires a careful comparison in the future and is beyond the scope of this paper.

## 2. The dynamic linear model

Throughout this paper we will use the notation as in [3].

### 2.1. Definition of the dynamic linear model

Let $Y_t \in \mathbb{R}$ be the observation at time $t \in [0, 1 \dots T\}$. Then for each time $t$, the dynamic linear model is defined by

$$\text{Observation equation: } Y_t = F_t'\theta_t + v_t, \quad v_t \sim N(0, V_t),$$
$$\text{System equation: } \theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t),$$

where $F_t \in \mathbb{R}^n, G_t \in \mathbb{R}^{n\times n}$ are known design matrices describing the deterministic part of the observation process and of the system evolution. Both processes are disturbed by the Gaussian noise terms $v(t) \in \mathbb{R}$ (observation variance) and $w(t) \in \mathbb{R}^n$ (evolution variance), which are assumed to be mutually independent with variances $V_t \in \mathbb{R}$, $W_t \in \mathbb{R}^{n\times n}$.

The model is initialized by a known prior for the initial state vector $\theta_0 \in \mathbb{R}^n$, usually taken to be

$$(\theta_0|D_0) \sim N(m_0, C_0),$$

where generally $D_t = \{Y_t, \dots, Y_0\} = \{Y_t, D_{t-1}\}$ represents the information set at time $t$, such that $D_0$ represents the initial information.

The dynamic linear model with design matrices $F_t, G_t$ and variances $V_t, W_t$ may symbolically be written as $M_t = \{F, G, V, W\}_t$.

### 2.2. Estimation of the state vector $\theta_t$

The following updating equations are used in estimating $\theta_t$ (see also [3]):

(a) Given posterior information at time $t-1$

$$(\theta_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1})$$

we arrive at the following.

(b) Prior at time $t$

$$(\theta_t|D_{t-1}) \sim N(a_t, R_t),$$

where $a_t = G_t m_{t-1}$ $R_t = G_t C_{t-1} G_t' + W_t$. Next we can forecast $Y_t$, by using information up to time $t-1$.

(c) One-step forecast

$$(Y_t|D_{t-1}) \sim N(f_t, Q_t),$$

where $f_t = F_t' a_t, Q_t = F_t' R_t F_t + V_t$. Eventually we obtain the following.

(d) Posterior at time $t$

$$(\theta_t|D_t) \sim N(m_t, C_t),$$

where $m_t = a_t + A_t e_t, C_t = R_t - A_t A_t' Q_t$ with $A_t = R_t F_t Q_t^{-1}$, $e_t = Y_t - f_t$.

## 3. A multi-process model for the on-line monitoring problem

Combinations of different filters are called multi-process models. Let $\mathscr{A}$ be some index set and for $\alpha \in \mathscr{A}$ let $M_t(\alpha)$ be the model corresponding to $\alpha$ (for some $F, G, V, W$ depending on $t$ and $\alpha$).

In the simplest case there is some fixed (though maybe unknown) $\alpha$ such that the model $M_t(\alpha)$ holds for all $t$, and this is what turns out to be general enough to handle the on-line problem.

For the estimation of $\alpha$ we use Bayes theory. Given an initial prior $p(\alpha|D_0)$, inferences about $\alpha$ can be drawn sequentially by $p(\alpha|D_t) \propto p\,(\alpha|D_{t-1})$ $p(Y_t|\alpha, D_{t-1})$.

Our multi-process model for the on-line monitoring problem – including multiple change-points – is based on the dynamic change-point model developed in [2], covering situations with at most one change-point. In the following we give a brief description of the latter model.

### 3.1. The dynamic change-point model

The structural component model of [2] describes a system without a change-point by a simple random walk. Change-points are incorporated by a "switch", which adds at some fixed but unknown time $\tau$ a (possibly noisy) drift to the system equation. Thus for each $\tau \in \{0, 1, \dots, T\}$ the observation and system equations are:

Observation equation: $Y_t = \mu_t + v_t, \quad v_t \sim N(0, \sigma_y^2)$,

System equation: $\mu_t = \mu_{t-1} + z_t^{(\tau)}\beta_{t-1} + w_{1t}, \quad w_{1t} \sim N(0, \sigma_\mu^2)$,

$\beta_t = \beta_{t-1} + w_{2t}, \quad w_{2t} \sim N(0, \sigma_\beta^2)$,

where $z_t^{(\tau)}$ is an indicator variable with

$$z_t^{(\tau)} = \begin{cases} 0 & t < \tau, \\ 1 & t \geqslant \tau. \end{cases}$$

We shall use the following notation: the "0-filter" refers to a filter with $z_t^{(\tau)} = 0$ and the "1-filter" refers to a filter with $z_t^{(\tau)} = 1$.

Every $\tau \in \{0, 1, \dots, T\}$ defines a different model. The collection of all these single-process models labeled by $\tau$ is called the multi-process model $M_t(\tau)$. In matrix notation:

$$Y_t = [1 \quad 0]_{F'}\theta_t + v_t,$$

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}_{\theta_t} = \begin{bmatrix} 1 & z_t(\tau) \\ 0 & 1 \end{bmatrix}_{G_t^\tau} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix}_{\theta_{t-1}} + \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}_{w_t},$$

with

$$\mathrm{Var}(v_t) =: V_t, \ \mathrm{Var}(w_t) =: W_t.$$

We discuss the problem of choosing $V_t$ and $W_t$ in Section 4.2. The updating equations given a change-point $\tau = j$ are described in Section 2.2

### 3.2. The estimation of change-points

The posterior distributions of the change-points $P(\tau = j|D_t)$ $j = [1, \dots, T]$, can be calculated using a Bayes approach from Bayes with:

$$P(\tau = j|D_t) \propto P(Y_t|D_{t-1}, \tau = j)P(\tau = j|D_{t-1}).$$

These probabilities must be initialized. If $\pi$ denotes the probability that a change-point occurs until time $T$, a reasonable initial prior is the uniform prior:

$$P(\tau = j|D_0) = \frac{\pi}{T}, \quad j = 1 \dots T.$$

For the estimation of $\tau$ it is only necessary to consider models up to time $t$, since all conditional models with $\tau \geqslant t+1$ are identical:

$$P(Y_t|D_{t-1}, \tau = j) = P(Y_t|D_{t-1}, \tau = t+1), \quad j \geqslant t+1.$$

Hence the posterior distribution of the change-point $\tau$ at time $t$ is given by

$$P(\tau = j|D_t) = \begin{cases} c_t\Phi(y_t; f_t^j, Q_t^j) \cdot P(\tau = j|D_{t-1}), & j \leqslant t, \\ c_t\Phi(y_t; f_t^{t+1}, Q_t^{t+1}) \cdot P(\tau = j|D_{t-1}), & t < j \leqslant T, \end{cases}$$

$$P(\tau > T|D_t) = c_t\Phi(y_t; f_t^{t+1}, Q_t^{t+1}) \cdot P(\tau > T|D_{t-1}),$$

where $f_t^j = E(Y_t|D_{t-1}, \tau = j)$ and $Q_t^j = \mathrm{Var}(Y_t|D_{t-1}, \tau = j)$ are the mean and variance of the one-step forecast density (see Section 2) and $\Phi$ is the density of the normal distribution, $c_t$ being the normalization constant.

The dynamic linear change-point model seems to be an appropriate model, which allows to detect on-line deviations from an assumed course of a monitored variable. But there exist still some unsolved problems:

Outliers can have an important influence on the probability of a change-point.

Long observation periods entail the need for handling many models simultaneously such that the algorithm becomes too slow for real time applications.

The original model only allows one to detect at most one change-point during the observation period.

The variances $V_t$ and $W_t$ are in many practically important cases unknown.

The next chapter shows how these problems can be solved.

## 4. Towards an on-line monitoring alert system

### 4.1. Introduction of a time window for the 1-filters

The computational time increases rapidly with the increasing number of *1-filters* to be processed such that the speed may easily drop below the limit for real time applications. To overcome this problem, we introduce a window $[t-b, t]$ for these 1-filters, with some positive $b$ depending on the computational power and the specific problem. Then the probabilities for the change-points are

$$P(\tau = j|D_t) = \begin{cases} c_t\Phi(y_t; f_t^j, Q_t^j) \cdot P(\tau = j|D_{t-1}), & t-b \leqslant j \leqslant t, \\ c_t\Phi(y_t; f_t^{t+1}, Q_t^{t+1}) \cdot P(\tau = j|D_{t-1}), & t < j \leqslant T, \\ 0, & j < t-b, \end{cases}$$

$$P(\tau > T|D_t) = c_t\Phi(y_t; f_t^{t+1}, Q_t^{t+1}) \cdot P(\tau > T|D_{t-1}).$$

Only the 0-*filter* and the $t-b+1$ 1-filters are considered in the calculations of the change-point probabilities. One obvious advantage is that the time needed to perform the various calculations stays more or less constant, especially not growing with time. An additional advantage is that the model is now able to deal with more than one change-point. Since a change-point before time $t-b$ is no longer respected, we estimate the posterior distribution of the actual change-point using information only from within this time window. However, the window $t-b$ is dynamic. One 1-filter is added for the new observation and in the same moment we drop the 1-filter for observation $t-b$. Hence, in moving the window over time we are able to detect sequential change-points.

### 4.2. Hierarchical multi-process models

Let $P(M_t^{(\alpha)}|D_t)$ be the probability that the model $M_t^{(\alpha)}$, for some $\alpha \in \mathscr{A}$, holds at time $t$. Then we define a hierarchical model by the probability

$$P(M_t^{(\alpha)}, M_t^{(\beta)}|D_t) = P(M_t^{(\alpha)}|M_t^{(\beta)}, D_t)P(M_t^{(\beta)}|D_t),$$

where $\beta \in \mathscr{B}$, and $\mathscr{A}, \mathscr{B}$ are disjoint parameter sets.

If one is interested in marginal probabilities one may calculate them via

$$P(M_t^{(\alpha)}|D_t) = \sum_{i \in \mathscr{B}} P(M_t^{(\alpha)}|M_t^{(\beta_i)}, D_t)P(M_t^{(\beta_i)}|D_t).$$

This definition should not be confused with a multi-process model of class II, in which one will not distinguish between $\mathscr{A}$ and $\mathscr{B}$. A hierarchical model is the combination of two or more multi-process models of class I. So one is able to follow a decision tree within the set of different filters. We will use it to build an estimation procedure for the unknown $V_t$ and $W_t$, as well as for the outlier detection.

Before we propose the estimation procedures for the unknown variances, we discuss some basic considerations. Up to now we have not distinguished between different filters and their variances. However, this will become important when we are going to estimate these variances on-line. The fundamental approach to the on-line monitoring problem using the dynamic linear change-point model is that the new observation $Y_t$ is explained by two types of models (the 0- and 1-filters). A change-point is detected when the 1-filters are better in predicting $Y_t$ than the 0-filter.

Since the system equations of the 0- and 1-filters are different ($\mu_t = \mu_{t-1} + w_t^{(0)}$ for the 0-filter and $\mu_t = \mu_{t-1} + \beta_{t-1} + w_t^{(\tau)}$ for the 1-filters) one will have different evolution variances $W_t^{(0)}, W_t^{(\tau)}, \tau = 1, \dots, t$. The only difference between 0- and 1-filters is the slope parameter $\beta$. Hence, in adding a slope parameter to the system equation a part of the evolution variance estimated for the 0-filter, is now explained by the slope itself, and therefore $W_t^{(\tau)} \leqslant W_t^{(0)}$. However, the observation variance $V_t$, which has the interpretation of measurement error, is identical for both models, because the observation equations are identical.

These considerations lead to the following estimation concept. Since we are not able to find a analytic estimation procedure in terms of a single multi-process filter, to estimate $V_t$ and $W_t$ simultaneously we have to estimate these variances separately from the estimation of a change-point. Therefore we will introduce, for estimation of the unknown variances, a new multi-process filter consisting of the 0-filter and the 1-filter (with $\tau = 1$). This leads to what we call a *hierarchical* on-line estimation procedure.

### 4.3. *On-line estimation of the unknown variance $V_t$*

To estimate the unknown observational variance $V_t$ we adapted a conjugate sequential updating procedure described in West, Harrison (1989, 118ff). Since $V_t$ becomes now a random quantity the normal distribution changes into a $t$-distribution and we will obtain the following system:

$$\text{Observation equation: } Y_t = F'\theta_t + v_t, \quad v_t \sim N(0, V_t),$$
$$\text{System equation: } \theta_t = G_t^j\theta_{t-1} + w_t, \quad w_t \sim T_{n_{t-1}}(0, W_t), j \in \{T+1, 1\},$$

where $T_{n_{t-1}}(\mu, \sigma^2)$ denotes the noncentral $T$-distribution with mean $\mu$, variance $\sigma^2$ and $n_{t-1}$ degrees of freedom. The expression $j \in \{T+1, 1\}$ indicates two filters, one for the 0-filter and one for the 1-filter, that started from the beginning. The updating equations will now take the form:

(a) Posterior at $t-1$: $(\theta_{t-1}|D_{t-1}, j) \sim T_{n_{t-1}}(m_{t-1}, C_{t-1})$.
(b) Prior at $t$: $(\theta_t|D_{t-1}, j) \sim T_{n_{t-1}}(a_t, R_t), a_t^j = G_t^j m_{t-1}^j, R_t^j = G_t^j C_{t-1}^j G_t^{\prime j} + W_t$.
(c) one-step forecast: $(Y_t|D_{t-1}, j) \sim T_{n_{t-1}}(f_t, Q_t), f_t^j = F'a_t^j, Q_t^j = S_{t-1}^j + F'R_t^j F$.
(d) Posterior at t: $(\theta_t|D_t, j) \sim T_{n_{t-1}}(m_t, C_t), m_t^j = a_t^j + A_t^j e_t^j, C_t^j = S_t^j/S_{t-1}^j[R_t^j - A_t^j A_t^{\prime j} Q_t^j], S_t^j = d_t/n_t,$ where $n_t = n_{t-1} + 1, d_t = d_{t-1} + S_{t-1}^j e_t^{j2}/Q_t^j, A_t = R_t^j F/Q_t^j$ and $j \in \{T+1, 1\}$.

Under the assumption that the estimated variances $V_t^{(0)}$ and $V_t^{(\tau=1)}$ are now known quantities, we can combine the two filters in a multi-process model and use this to get an estimate of $V_t$ simultaneously.Using Bayes we get

$$P(V_t^j|D_t) \propto P(Y_t|V_t^j, D_{t-1})P(V_t^j|D_{t-1}), \quad j \in \{T+1, 1\}$$

and we can get an estimate of $V_t$ by

$$\widehat{V}_t = \sum_j V^j P(V_t^j|D_t).$$

As initial probabilities $P(V_t^j|D_0)$ one can use the probabilities $\pi, 1-\pi$, which were used to initialize $P(j|D_0)$ in the change-point estimation procedure. The single estimates of $V_t^j$ will be passed to the hierarchical change-point model.

### 4.4. On-line estimation of the unknown variance $W_t$

Similar to Section 4.3 we will build an estimation procedure to calculate $W_t$. In a first step we transform the problem of estimating $W_t$ to a problem where we have to estimate a discounted variance. As proposed by [3] we introduce a discounting factor $\delta_{t-1}$, with $0 < \delta_{t-1} \leqslant 1$. By definition we can set

$$W_t = P_t(1 - \delta_{t-1})/\delta_{t-1}$$

with

$$P_t = G_t C_{t-1} G_t'.$$

One advantage is now that in contrast to $W_t$, $\delta_{t-1}$ is scale free. Furthermore $\delta_{t-1}$ is related in the limit to the signal to noise ratio $r = W_t/V_t = (1 - \delta_{t-1})^2/\delta_{t-1}$. In the literature values like $\delta = 0.7$ or 0.9 are chosen to be fixed and the usual updating equations are used to estimate the state vector $\theta_t$. Hence, a possible strategy could be to analyze several data sets with defined change-points and to look for the best value of $\delta$. But this would not be an on-line estimation of the evolution variance $W_t$. Another possibility is to estimate the unknown discounting factor $\delta_{t-1}$ similarly to the observation variance $V_t$. Our proposal is to do the following:

As mentioned at the beginning of this chapter, we need a $\delta_{t-1}$ for the 0-filter and the 1-filter. So we have to build two different multi-process models. First let $j \in \{T+1, 1\}$. Then:

(a) choose a discrete set $[\delta_{t-1}^1, \delta_{t-1}^2, \ldots, \delta_{t-1}^k]$ of values for $\delta_{t-1}$ ($k$ appropriately chosen);

(b) calculate at each step the probabilities of $\delta_{t-1}$ using

$$P(\delta_{t-1}^j|D_{t-1}) \propto P(Y_t|\delta_{t-1}^j, D_{t-2}, j)P(\delta_{t-1}^j|D_{t-2}, j),$$

One may estimate $\delta_{t-1}^j$ using

$$\hat{\delta}^j_{t-1} = \sum_{i=1}^{k} \delta_{t-1} P(\delta^i_{t-1} | D_t, j).$$

It seems to be natural to use the uniform distribution $P(\delta_i|D_0) = 1/k, i = 1, \ldots, k$ for the initial probabilities. This method appears to be a good estimation strategy for the unknown $\delta^j_{t-1}$. Once again the estimated parameters are passed to the hierarchical change-point model.

### 4.5. Outliers

To detect outliers we use the ideas of [1]. An outlier can be interpreted as a sudden perturbation of the observation equation. To include this possibility we could enlarge the multi-process model by an extra filter for outliers (which we call "$N$-filter"), which is exactly the 0-filter with an enlarged observation variance. Since, $V_t$ and $W_t$ are estimated on-line this will not work. Instead of including the extra $N$-filter in the change-point model, we introduce an extra multi-process model for outliers. This model will become the first level of our hierarchical multi-process model. Let

$$M_t(\kappa)\colon Y_t = F'\theta_t + v_t, \quad v_t \sim N(0, \kappa V),$$
$$\theta_t = G_t^{(T+1)}\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t),$$

where $\kappa$ is chosen sufficiently large, say $\kappa = 100$. Now we may estimate the probability of an outlier by

$$P(\text{Outlier}|D_t) \propto P(Y_t|\text{Outlier}, D_{t-1})P(\text{Outlier}|D_{t-1}).$$

### 4.6. Initialization

We now show how the prior distributions of the state vectors should be specified. We initialize the 0-filter and 1-filter similar to [2] with a data driven prior. The variances $C_0^{(\tau)}$ follow hereby a diffuse prior.

The 0-filter $(\theta_0|D_0, T+1) \sim N(m_0^{(T+1)}, C_0^{(T+1)})$ is initialized by

$$m_0^{(T+1)} = (Y_1, 0)' \quad \text{and} \quad C_0^{(T+1)} = \begin{pmatrix} V_0 + W_0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Prior information of the 1-filters $(\theta_0|D_0, \tau \leqslant t) \sim N(m_0^{(\tau\leqslant t)}, C_0^{(\tau\leqslant t)})$ is recursively defined by

$$m_0^{(\tau=t)} = (m_{t-1,1}^{(\tau=t-1)}, Y_{t-1} - m_{t-1,1}^{(\tau=t-1)})',$$
$$C_0^{(\tau=t)} = \begin{pmatrix} C_0^{(\tau=t-1)} & -C_0^{(\tau=t-1)} \\ -C_0^{(\tau=t-1)} & C_0^{(\tau=t-1)} + V_t + W_t \end{pmatrix}.$$

Furthermore, we have to specify starting values for the variances $V_t$, $W_t$ and we have to choose a discrete set for the discounting factor $\delta$. Since we are going to estimate all parameters on-line we need a good guess for the signal to noise ratio $r$ of the underlying process. Otherwise the estimation procedure will not converge to the true values.

For the approximation of $r$ we will use the first $2k$ observations (chose $k$ appropriately) and we define the following quantities:

$$a = \mathrm{Var}(Y_i - Y_{i-1}), \quad i = 1, \ldots, 2k,$$

$$b = \mathrm{Var}(Y_i + Y_{2k-i+1}), \quad i = 1, \ldots, k.$$

Then

$$r = 2\frac{a - b}{(2k+1)b - a}.$$

With this we are able to choose $V$ and $W$ such that $r = W/V$. Furthermore, we can now choose a discrete set for $\delta$. Since $r = (1-\delta)^2/\delta$ it is convenient to

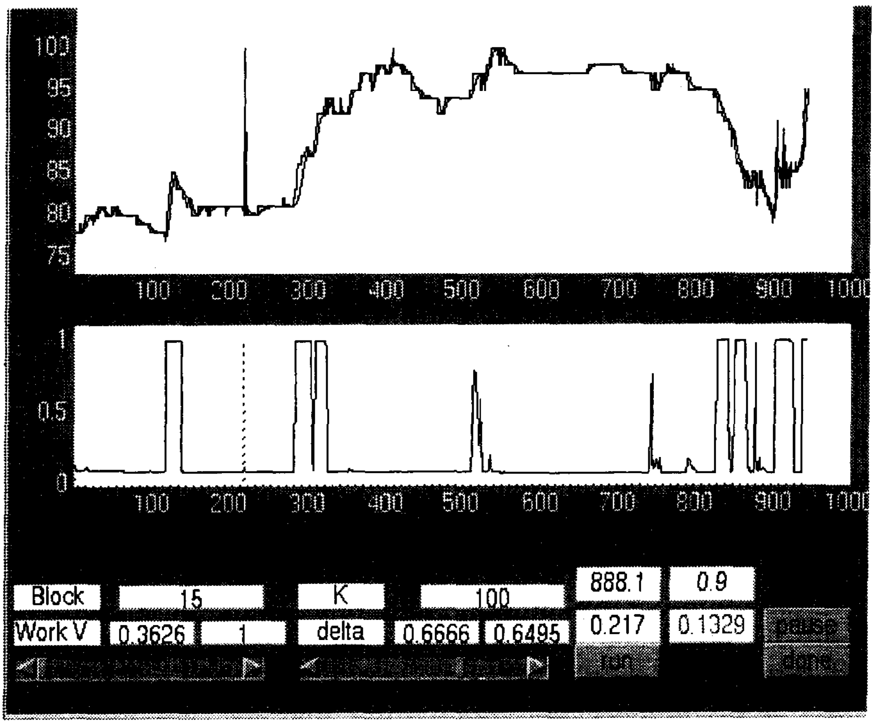


Fig. 1. We see from the figure, that at observation 120 an alert is given. This coincides with the beginning of the first skin cut. At 250 we have introduced an outlier, which was detected by the $N$-filter. At 285 the operation starts. Since the patient reacted to this, the anesthesiologist had to intervene. The weak change-point at 506 was in this period of external stabilization. From 540 to 752 we have a stable phase. At 752 one observes a weak change-point. This marks the end of the operation and the anesthesiologist initiated the wakeup phase. At 821 the patient awoke.

choose a discrete interval of $\delta$ about $r$. This interval can then be updated as new observations are made using the same approximation as before.

## 5. Example

The following data (Fig. 1) are the ECG measurements, taken every five seconds, from a patient undergoing a skin transplantation. Monitoring started when the first steps in preparing the patient were finished and anesthesia was completed.

The first window shows the ECG measurements with the filtered values of the 0-filter. The second window displays the estimated cumulative probability that a change-point has occurred within the observation window. Furthermore we display the probability of an outlier at the current timepoint.

## 6. Conclusions

We have presented an algorithm based on multi-process Kalman filters, which allows one to estimate the probability that a change-point has occurred in a noisy time series. The algorithm is stable against outliers, adaptive and fast enough for on-line applications. In situations where one has good control of the system and observation equation, the signal to noise ratio, independence gaussian distribution etc., the well-known optimality results for the Kalman-filter apply. We shall discuss extensions, alternative algorithms and detailed methods for validation elsewhere.

## Acknowledgements

For helpful comments we thank especially Prof. Dr. Ludwig Fahrmeir, Prof. Dr. Albrecht Neiss and Dr. Stefan Wagenpfeil.

## References

[1] A.F.M. Smith, M. West, Monitoring renal transplants: An application of the multi-process Kalman filter, Biometrics 39 (1983) 867–878.
[2] J. Whittaker, Frühwirth-Schnatter, A dynamic change-point model for detecting the onset of the growth in bacteriological infections, Appl. Statist. 43 (4) (1994).
[3] M. West, P.J. Harrison, Bayesian Forecasting and Dynamical Models, Springer, New York, 1989.
[4] G. Kitagawa, Monte Carlo smoother for non-gaussian non-linear state space models, J. Comput. Graphical Stat. 5 (1996) 1–25.
[5] G. Kitagawa, N. Matsumoto, Detection of coseismic changes of underground water level, Jasa 91 (1996) 521–528.
[6] S. Kim, N. Shephard, S. Chib, Stochastic volatility: Likelihood inference and comparison with arch models, Technical Report, 1996.